

ANALYSIS OF COMPLEX ALLOZYME POLYMORPHISMS IN A BARLEY POPULATION¹

B. S. WEIR², R. W. ALLARD³ AND A. L. KAHLER³

University of California, Davis, California 95616

Manuscript received December 27, 1971

Revised copy received June 26, 1972

ABSTRACT

Genotypes of 68,230 individuals taken from 10 generations (F_4 - F_6 , F_{14} - F_{17} , F_{24} - F_{26}) of an experimental population of barley were determined for four esterase loci. The results show that frequencies of gametic ditypes changed significantly over generations and that striking gametic phase disequilibrium developed within a few generations for each of the six pairwise combinations of loci which were monitored. The complex behavior of these four enzyme loci in the population is attributed to interactions between selection and restriction of recombination resulting from the effects of linkage and/or inbreeding.

THIS is the third in a series of papers dealing with the analysis of the frequency of allozymes at four esterase loci in Composite Cross V (CCV), an experimental population of barley (*Hordeum vulgare* L.) developed from intercrosses among 30 barley varieties from various parts of the world. The first paper (KAHLER and ALLARD 1970, referred to below as I) described the electrophoretic techniques used to detect the allozymes and presented the results of formal genetic studies of these esterase loci. In the second paper (ALLARD, KAHLER and WEIR 1972, referred to below as II), the history of CCV was described and the single-locus frequency data were analyzed. The results showed that mutation, migration, and genetic drift had at most trivial effects and that selection and mating system were the major forces responsible for the patterns of genetic change which occurred in the population. In this paper we extend the analysis of data for CCV to pairs of loci. It will be shown that interactions between selection and restriction of recombination due to linkage and/or inbreeding play a predominant role in the behavior of these four enzyme loci in the population.

TWO LOCUS DATA

The materials on which this study was conducted and the methods employed in obtaining the data are described in detail in I and II and hence will not be presented here. The basic data consisted of the multilocus esterase genotypes of

¹ Supported in part by grants from the National Institutes of Health (GM-10476) and the National Science Foundation (GB-13213).

² Department of Agronomy and Range Science. Present address: Department of Mathematics, Massey University, Palmerston North, New Zealand.

³ Department of Genetics.

68,230 individuals from 10 generations of CCV: three early generations (F_4 , F_5 , F_6), four intermediate generations (F_{14} , F_{15} , F_{16} , F_{17}) and three late generations (F_{24} , F_{25} , F_{26}). The four loci monitored, EA, EB, EC and ED, here designated A, B, C and D, are represented by four, three, three and four alleles, respectively, in CCV (I, II). Thus, considering each locus individually, there are 6 genotypes for Loci B and C and 10 for Loci A and D; considering the loci pairwise there are 36 genotypes for the B-C combination, 60 for the A-B, A-C, B-D and C-D combinations, and 100 for the A-D combination. A few of these genotypes cannot be distinguished phenotypically due to recessiveness of a null allele at the D locus, or to specific interactions between the $A^{2.6}$ and $B^{2.7}$ alleles (I, II). Nevertheless, the number of genotypes that can be identified is large and, hence, for ease of presentation we seek to reduce these numbers in some way. There is another reason for such reduction. CCV practices 99.43% self fertilization (II) so that certain heterozygotes (especially double heterozygotes representing rare alleles) are infrequent in the population and such heterozygotes are sometimes not present even in the large samples studied. This causes problems in analyses which attempt to take all genotypes into account. We use the two following methods of reducing and simplifying the data: (1) consider homozygous classes only, thus reducing the number of genotypes from 36 to 9 for the B-C combination, from 60 to 12 for A-B, A-C, B-D and C-D combinations, and from 100 to 16 for the A-D combination; (2) combine the data into two allelic classes per locus, one consisting of the most frequent allele and the other consisting of all other alleles combined.

ANALYSIS OF HOMOZYGOUS CLASSES

The data for the six possible pairwise comparisons of homozygotes in each of 10 generations of CCV are too extensive to be given in their entirety. However results are similar within comparisons between pairs of the three tightly-linked loci A, B and C ($B \leftarrow 0.0023 \pm 0.0007 \rightarrow A \leftarrow 0.0048 \pm 0.0008 \rightarrow C$), within comparisons of each of these loci with the non-linked locus D, and also within the early, intermediate and late generations. Consequently data will be given for only one early, one intermediate and one late generation for one pair of linked loci and one pair of unlinked loci, since these partial data suffice to bring out the main features.

It can be deduced from the genotypes of the parents of CCV (see Table 11) that alleles at pairs of loci came into the population approximately in random association, i.e., two-locus genotypic frequencies corresponded fairly closely to those calculated as products of appropriate single-locus marginal frequencies in the initial (F_2) generation of the population. This situation still prevailed in the early (F_4 , F_5 , F_6) generations for the three pairs of unlinked loci as illustrated in Table 1, which gives relative deviations of two-locus genotypic numbers from numbers (rounded to nearest integer) predicted from single-locus frequencies for loci B-D. In generation 6 departures from expected values are small and non-significant by a chi square goodness-of-fit test. However, by generation 17 there is clear evidence that certain alleles at the two loci (e.g. $B^{1.6}$ and $D^{6.4}$) interact

TABLE 1

Relative deviations [(observed two-locus numbers minus the product of one-locus numbers/ N) \times 1000] for loci B and D in generations 6, 17, and 26 of CCV

	Locus B [†]	Locus D [†]			
		11	22	33	44
Generation 6	11	— 5	— 1	+ 6	+ 1
$N = 1006$	22	+ 4	+ 2	— 5	— 1
$\chi^2 = 5.1$ (NS)	33	+ 1	— 1	— 1	— 1
Generation 17	11	+ 8	+ 2	— 8	— 1
$N = 2461$	22	—12	— 2	+ 7	+ 8
$\chi^2 = 49.5$	33	+ 3	+ 1	+ 2	— 7
Generation 26	11	+45	— 9	—25	—10
$N = 3083$	22	—48	+ 9	+20	+14
$\chi^2 = 240.8$	33	— 1	— 1	+ 3	— 4

[†] The allelic designations for locus B are B^{1.6} = 1, B^{2.7} = 2, B^{3.9} = 3 and for locus D they are D^{6.4} = 1, D^{6.5} = 2, D^{6.6} = 3 and D^N = 4.

favorably and that others (e.g. B^{2.7} and D^{6.4}) interact unfavorably with each other in their homozygous combinations. By generation 26 departures have become very large. Inspection of this table shows that each of the four alleles at the D locus interacts favorably in at least one of its homozygous combinations with the three alleles at the B locus, and unfavorably in at least one combination. Results are closely similar for the two other comparisons of unlinked loci (A-D and C-D).

Table 2 gives results for the tightly-linked pair of loci, B and C. In this case departures from expected values are significant by generation 6 and it can be

TABLE 2

Relative deviations [(observed two-locus numbers minus the product of one-locus numbers/ N) \times 1000] for loci B and C in generations 6, 17, and 26 of CCV

	Locus B [†]	Locus C [†]		
		11	22	33
Generation 6	11	+ 17	— 5	— 11
$N = 1006$	22	— 28	+ 1	+ 26
$\chi^2 = 124.9$	33	+ 10	+ 4	— 15
Generation 17	11	+ 26	—15	— 11
$N = 2461$	22	— 53	+18	+ 34
$\chi^2 = 1236.5$	33	+ 26	— 3	— 22
Generation 26	11	+125	—43	— 80
$N = 3083$	22	—160	+47	+115
$\chi^2 = 2187.8$	33	+ 35	— 3	— 32

[†] The allelic designations for locus B are B^{1.6} = 1, B^{2.7} = 2, B^{3.9} = 3 and for locus C they are C^{4.4} = 1, C^{4.9} = 2, and C^{5.4} = 3.

seen that each of the three alleles at the B locus interacts favorably in at least one homozygous combination with alleles at the C locus, and unfavorably in other homozygous combinations. Departures from expectations based on marginal frequencies have become very large by generation 17 and they are larger still by generation 26. Results were similar for the two other pairs of linked loci (A-B and B-C).

The changes which occurred in CCV for all six pairs of loci were thus from random associations of alleles in the initial generation to non-random associations in later generations, a pattern of change which is at variance with selective neutrality. The directions of the departures from randomness which developed indicate further that the selection which occurred featured complex epistatic interactions between alleles at different loci.

ANALYSIS IN TERMS OF TWO ALLELES PER LOCUS

In this form of analysis we reduce the data to two allelic classes, denoting the class consisting of the single most frequent allele by subscript 1 and the other class, consisting of all other alleles combined, by subscript 2. The correspondence between these new "alleles" and the actual alleles in the population (I, II) are as follows:

$$\begin{array}{ll} A_1 = A^{1.8} & B_1 = B^{2.7} \\ A_2 = A^{0.2} + A^{1.0} + A^{2.6} & B_2 = B^{1.6} + B^{3.9} \\ C_1 = C^{5.4} & D_1 = D^{6.4} \\ C_2 = C^{4.4} + C^{4.9} & D_2 = D^{6.5} + D^{6.6} + D^N. \end{array}$$

In some cases considerable information is lost by reducing the data in this way. Thus, for example, combining alleles $C^{4.4}$ and $C^{4.9}$ into a single "new" allele leads to underestimation of the interaction between loci B and C because these two C-locus alleles interact in opposite directions in their homozygous combinations with B-locus homozygotes (Table 1).

TABLE 3

Two-locus relative genotypic frequencies for locus pair A-B

Generation	Number of plants in sample	Genotypes									χ^2	(d.f.)*
		1111	1122	2211	2222	1112	1211	1222	2212	1212		
4	1234	.4076	.0584	.4206	.0162	.0307	.0584	.0008	.0024	.0049	60.47	(4)
5	1486	.4367	.0525	.4367	.0222	.0081	.0390	0	.0014	.0034	23.70	(3)
6	1006	.4294	.0786	.4284	.0348	0	.0278	.0010	0	0	17.09	(1)
14	1651	.3895	.0896	.4809	.0030	.0164	.0182	0	.0012	.0012	173.41	(3)
15	2843	.3356	.1330	.4625	.0018	.0405	.0128	.0018	.0004	.0026	538.43	(4)
16	2369	.4432	.0646	.4458	.0241	.0106	.0110	0	.0004	.0003	61.83	(3)
17	2461	.4917	.0678	.4153	.0199	.0008	.0045	0	0	0	42.90	(1)
24	4396	.3710	.2109	.3776	.0021	.0150	.0177	.0005	.0002	.0050	796.90	(3)
25	3967	.3708	.2180	.3864	.0018	.0081	.0098	.0025	.0003	.0023	735.54	(3)
26	3083	.3477	.2520	.3821	.0019	.0042	.0078	.0020	.0003	.0020	654.24	(3)

* All χ^2 values significant at $P = .001$.

TABLE 4

Two-locus relative genotypic frequencies for locus pair A-C

Generation	Number of plants in sample	Genotypes										χ^2	(d.f.)*
		1111	1122	2211	2222	1112	1211	1222	2212	1212			
4	1234	.3168	.1297	.3039	.1159	.0494	.0356	.0065	.0203	.0219	76.80	(4)	
5	1486	.2874	.1837	.3183	.1339	.0262	.0162	.0047	.0081	.0215	278.04	(4)	
6	1006	.3101	.1958	.3042	.1541	.0020	.0159	.0020	.0050	.0109	21.21	(3)	
14	1651	.3410	.1133	.3701	.0878	.0412	.0170	.0012	.0272	.0012	15.40	(3)	
15	2843	.2726	.2279	.3394	.1224	.0085	.0137	.0025	.0028	.0102	128.68	(3)	
16	2369	.2672	.2356	.3377	.1245	.0156	.0030	.0004	.0080	.0080	119.23	(3)	
17	2461	.2926	.2662	.3170	.1154	.0016	.0028	0	.0028	.0016	113.77	(3)	
24	4396	.2543	.2807	.2771	.0958	.0619	.0139	.0048	.0070	.0045	430.45	(4)	
25	3967	.2059	.3857	.2637	.1230	.0053	.0058	.0050	.0018	.0038	422.08	(3)	
26	3083	.1768	.4220	.2809	.1015	.0052	.0036	.0036	.0019	.0045	569.61	(3)	

* All χ^2 values significant at $P = .001$.

In general we write the relative frequency of a quantity x as $f(x)$. For the two-locus genotypes then we have such frequencies as $f(A_1A_1B_2B_2)$ and $f(A_1A_1B_1B_2)$. On some occasions we need to take notice of which genes were received from the same gamete. For example for genotypes formed by the union of gametes A_iB_j and A_kB_l (i, j, k and l are not necessarily different) we write the frequency as $f(A_iB_j, A_kB_l)$ or even just as $f(ij, kl)$ when it is clear to which locus pair we are referring. Note that $f(ij, kl) = f(kl, ij)$, so that heterozygote frequencies may be written as twice either of two equal quantities. For example we can write $2f(A_1B_1, A_1B_2)$ for $f(A_1B_1, A_1B_2) + f(A_1B_2, A_1B_1)$ where both expressions equal $f(A_1A_1B_1B_2)$. The coupling and repulsion double heterozygotes can not be distinguished without progeny testing, and since this was impracticable, only the sum of these two classes is available. The generation time for all frequencies may be designated with a superscript.

TABLE 5

Two-locus relative genotypic frequencies for locus pair B-C

Gener- ation	Number of plants in sample	Genotypes										χ^2	(d.f.)*
		1111	1122	2211	2222	1112	1211	1222	2212	1212			
4	1234	.6240	.1953	.0292	.0365	.0673	.0032	.0203	.0097	.0145	124.17	(4)	
5	1486	.5935	.2719	.0256	.0451	.0471	.0027	.0054	.0040	.0047	59.56	(3)	
6	1006	.5845	.2843	.0457	.0676	.0169	0	0	.0010	0	32.08	(1)	
14	1651	.6964	.1386	.0267	.0539	.0535	.0049	.0097	.0121	.0042	216.24	(3)	
15	2843	.6201	.1868	.0035	.1298	.0130	.0021	.0362	.0032	.0053	948.74	(3)	
16	2369	.5804	.2980	.0270	.0587	.0215	.0004	.0038	.0030	.0072	149.64	(3)	
17	2461	.5920	.3133	.0195	.0683	.0061	.0008	0	0	0	157.56	(1)	
24	4396	.4941	.2318	.0482	.1383	.0405	.0030	.0111	.0268	.0062	617.79	(4)	
25	3967	.4726	.2889	.0010	.2180	.0055	.0018	.0069	.0033	.0020	1049.47	(3)	
26	3083	.4551	.2744	.0055	.2494	.0081	.0006	.0032	.0010	.0027	859.91	(3)	

* All χ^2 values significant at $P = .001$.

TABLE 6

Two-locus relative genotypic frequencies for locus pair A-D

Generation	Number of plants in sample	Genotypes								χ^2	(d.f.)	
		1111	1122	2211	2222	1112	1211	1222	2212			1212
5	1486	.2194	.2759	.1992	.2591	.0020	.0195	.0229	.0020	0	0.21	(3)
6	1006	.2704	.2316	.2336	.2286	.0060	.0159	.0119	.0010	.0010	3.24	(3)
16	2369	.2684	.2486	.2288	.2402	.0013	.0059	.0051	.0013	.0004	2.66	(3)
17	2461	.3072	.2511	.2251	.2044	.0020	.0033	.0012	.0057	0	10.29	(3)*
25	3946	.3876	.2051	.2243	.1591	.0050	.0091	.0040	.0043	.0015	23.26	(3)***
26	3083	.4619	.1369	.2523	.1242	.0052	.0068	.0039	.0078	.0010	47.87	(3)***

* χ^2 value significant $P = .05$, *** $P = .001$.

TABLE 7

Two-locus relative genotypic frequencies for locus pair B-D

Generation	Number of plants in sample	Genotypes									χ^2	(d.f.)
		1111	1122	2211	2222	1112	1211	1222	2212	1212		
5	1486	.4004	.5081	.0310	.0437	.0040	.0067	.0061	0	0	0.87	(2)
6	1006	.4642	.4145	.0557	.0576	.0070	0	0	.0010	0	0.55	(1)
16	2369	.4356	.4622	.0621	.0258	.0021	.0055	.0059	.0008	0	37.04	(2)***
17	2461	.4762	.4283	.0593	.0281	.0069	0	.0004	.0004	.0004	18.55	(1)***
25	3946	.4149	.3431	.1979	.0226	.0089	.0081	.0025	.0020	0	358.36	(3)***
26	3083	.4840	.2413	.2316	.0227	.0123	.0055	.0010	.0016	0	182.16	(3)***

*** χ^2 values significant at $P = .001$.

TABLE 8

Two-locus relative genotypic frequencies for locus pair C-D

Gener- ation	Number of plants in sample	Genotypes									χ^2	(d.f.)
		1111	1122	2211	2222	1112	1211	1222	2212	1212		
5	1486	.2483	.3708	.1615	.1602	.0027	.0282	.0269	.0007	.0007	14.92	(3)**
6	1006	.3181	.3081	.1889	.1590	.0040	.0129	.0050	.0040	0	4.60	(3)
16	2369	.3182	.2879	.1718	.1887	.0017	.0131	.0173	0	.0013	6.76	(3)
17	2461	.3454	.2609	.1861	.1938	.0061	.0041	.0020	.0016	0	18.26	(3)***
25	3946	.2525	.2174	.3616	.1480	.0048	.0068	.0028	.0048	.0013	123.36	(3)***
26	3083	.2997	.1524	.4139	.1087	.0091	.0075	.0039	.0045	.0003	71.45	(3)***

** χ^2 value significant at $P = .01$; *** $P = .001$.

Two-locus genotype frequencies are given in Tables 3 through 8. The tables for Loci A, B and C are similar, with double homozygotes being by far the most frequent category in all generations. Heterozygosity decreased rapidly in the early generations (F_4 , F_5 and F_6), as expected in an originally highly heterozygous population which practices more than 99% self fertilization. It is further expected (II) that the population will closely approach inbreeding equilibrium by

the sixth or seventh generation. Since no consistent changes occurred in the frequency of heterozygotes in the intermediate or late generations, this expectation also appears to have been realized in CCV.

For Loci A and B the A_1B_2/A_1B_2 homozygote (12 allelic combination) increased at the expense of all other double homozygotes. Similar changes took place for the A and C, and B and C locus pairs with the A_1C_2/A_1C_2 and B_2C_2/B_2C_2 homozygotes increasing in frequency. Thus the initially most frequent allelic combinations were not those most favored by environmental conditions at Davis, California.

Similar general comments can be made for the unlinked locus pairs A-D, B-D and C-D. The main difference is that the level of heterozygosity is lower for these locus pairs in the early generations and that the level changes little over generations. However, as discussed earlier (II), this is almost certainly due to underestimation of heterozygosity arising from inability to identify heterozygotes involving the D^N allele. This allele does not produce a band in starch gels and, since it is recessive to alleles which produce bands, its heterozygotes are wrongly classified as homozygotes.

A test for interactions between pairs of loci parallel to that used above for homozygous combinations can be made by comparing observed two-locus genotypic frequencies with expectations calculated from one-locus frequencies. One-locus frequencies can be obtained from the data in Tables 3-8 by appropriate summations of two-locus frequencies. For example, the frequency of A_1A_1 can be obtained from the A-B frequencies as: $f(A_1A_1) = f(A_1A_1B_1B_1) + f(A_1A_1B_1B_2) + f(A_1A_1B_2B_2)$. In more general notation, for the A locus:

$$f(A_iA_k) = \sum_{jl} f(A_iB_j, A_kB_l) = \sum_{jl} f(A_iC_j, A_kC_l) = \sum_{jl} f(A_iD_j, A_kD_l).$$

Treating the products of one-locus frequencies [such as $f(A_iA_k)$ and $f(B_jB_l)$] as the expected values of the observed two-locus frequencies [such as $f(A_iA_kB_jB_l)$], we form the chi-square goodness-of-fit statistics shown in the right hand columns of Tables 3-8. Chi-square values have four degrees of freedom (there are eight independent two-locus frequencies, from which four independent one-locus frequencies have been estimated) except where it was necessary to combine classes to give expected values larger than three. Exact expressions for the expected discrepancies between two-locus frequencies and the product of corresponding one-locus frequencies for neutral alleles are given by WEIR and COCKERHAM (in preparation). These discrepancies, while remaining non-zero, tend to be very small quantities which would certainly give non-significant chi-square values; also the χ^2 values are not expected to increase over generations with neutral alleles. The χ^2 values for the six pairwise comparisons between loci A, B, C, and D all increase and are consistently significant by the middle or late generations, thus providing evidence that these loci are under selection.

Although the χ^2 values reported in Tables 3-8 give a quantitative measure of departures from the neutral situation, the observed and relative deviations from their expected numbers give a more informative picture of the nature of the re-

organization of the two-locus zygotic arrays that has occurred over generations. Full data for observed and relative deviations are given for locus pair B-C in Table 9, where for convenience the deviations $f(B_i B_k C_j C_l) - f(B_i B_k) f(C_j C_l)$ have been rounded to the nearest integer. It can be seen that there is a consistent excess of $B_1 B_1 C_1 C_1$ and $B_2 B_2 C_2 C_2$ and a consistent deficit of $B_1 B_1 C_2 C_2$ and $B_2 B_2 C_1 C_1$ double homozygotes (compare with Table 2, which gives data for 3 alleles/locus). The deviations for the heterozygous classes are erratic due to sampling errors so that patterns are more difficult to identify. However, two of the singly heterozygous classes ($B_1 B_2 C_2 C_2, B_2 B_2 C_1 C_2$) and the double heterozygote ($B_1 B_2 C_1 C_2$) are in excess in most generations, while the other two singly heterozygous classes ($B_1 B_1 C_1 C_2, B_1 B_2 C_1 C_1$) are in deficiency in most generations. This is a drastically different pattern from that expected for neutral alleles (WEIR and COCKERHAM, in preparation). Only part of the results (observed and relative deviations from expected numbers in generations 6, 17 and 26) are given for the other pairs of loci (Table 10) because these partial results are adequate to establish that similar complicated epistatic interactions also occur in each of these cases.

Another way of characterizing two-locus behavior in CCV is in terms of changes in the frequencies of gametic ditypes over generations. Gametic ditype

TABLE 9

*Observed deviations of two-locus numbers from products of one-locus numbers for locus pair B-C in three early, four intermediate, and three late generations of CCV. Relative deviations [(observed deviation/N) × 1000] are in parentheses**

Generation	Genotypes								
	1111	1122	2211	2222	1112	1211	1222	2212	1212
4	+ 52 (+ 42)	- 35 (- 28)	- 25 (- 20)	+ 22 (+ 18)	- 17 (- 14)	- 27 (- 22)	+ 13 (+ 10)	+ 3 (- 2)	+ 14 (+ 11)
5	+ 39 (+ 26)	- 33 (- 22)	- 31 (- 21)	+ 31 (+ 21)	- 6 (- 4)	- 8 (- 5)	+ 2 (+ 1)	0 (0)	+ 6 (+ 4)
6	+ 26 (+ 26)	- 28 (- 28)	- 26 (- 26)	+ 28 (+ 28)	+ 1 (+ 1)	0 (0)	0 (0)	- 1 (- 1)	0 (0)
14	+ 82 (+ 50)	- 68 (- 41)	- 67 (- 41)	+ 58 (+ 35)	- 14 (- 8)	- 15 (- 9)	+ 10 (+ 6)	+ 9 (+ 5)	+ 5 (+ 3)
15	+ 304 (+ 107)	- 291 (- 102)	- 233 (- 82)	+ 232 (+ 82)	- 13 (- 5)	- 71 (- 25)	+ 59 (+ 21)	+ 1 (+ 0)	+ 12 (+ 4)
16	+ 79 (+ 33)	- 63 (- 27)	- 64 (- 27)	+ 63 (+ 27)	- 16 (- 7)	- 14 (- 6)	- 1 (- 0)	0 (0)	+ 16 (+ 7)
17	+ 83 (+ 34)	- 85 (- 34)	- 84 (- 34)	+ 86 (+ 35)	+ 1 (+ 0)	+ 1 (+ 0)	- 1 (- 0)	- 1 (- 0)	0 (0)
24	+ 335 (+ 76)	- 265 (- 60)	- 299 (- 68)	+ 250 (+ 57)	- 70 (- 16)	- 35 (- 8)	+ 15 (+ 3)	+ 49 (+ 11)	+ 20 (+ 4)
25	+ 428 (+ 108)	- 417 (- 105)	- 415 (- 105)	+ 412 (+ 104)	- 11 (- 3)	- 13 (- 3)	+ 5 (+ 1)	+ 3 (+ 1)	+ 8 (+ 2)
26	+ 354 (+ 115)	- 347 (- 112)	- 353 (- 114)	+ 354 (+ 115)	- 2 (- 1)	- 7 (- 2)	- 1 (- 0)	- 6 (- 2)	+ 8 (+ 3)

* Sample sizes given in Table 5.

TABLE 10
Observed deviations of two-locus numbers from products of one-locus numbers in one early, one intermediate and one late generation of CCV. Relative deviations $[(\text{observed deviation}/N) \times 1000]$ are in parentheses*

Locus pair	Generation	Genotypes									
		1111	1122	2211	2222	1112	1211	1222	2212	1212	
A-B	6	- 21(- 21)	+ 21(+ 21)	+ 18(+ 18)	- 18(- 18)	0(0)	+2(+2)	-2(-2)	0(0)	0(0)	
	17	- 47(- 19)	+ 46(+ 19)	+ 40(+ 16)	- 45(- 18)	+ 1(0)	+1(0)	-1(0)	- 1(0)	0(0)	
	26	-301(- 98)	+300(+ 97)	+304(+ 99)	-297(- 96)	+ 1(0)	-3(-1)	-3(-1)	- 7(-2)	+ 6(+ 2)	
A-C	6	- 10(- 10)	+ 17(+ 17)	+ 12(+ 12)	- 9(- 9)	- 7(- 7)	-2(-2)	-8(-8)	- 3(-3)	+10(+10)	
	17	-113(- 46)	+115(+ 47)	+123(+ 50)	-107(- 43)	- 2(-1)	-9(-4)	-9(-4)	-16(-6)	+18(+ 7)	
	26	-314(-102)	+320(+104)	+320(+104)	-312(-101)	- 6(-2)	-6(-2)	-8(-3)	- 8(-3)	+14(+ 4)	
B-C	6	+ 26(+ 26)	- 28(- 28)	- 26(- 26)	+ 28(+ 28)	+ 1(+ 1)	0(0)	0(0)	- 1(-1)	0(0)	
	17	+ 83(+ 34)	- 85(- 34)	- 84(- 34)	+ 86(+ 35)	+ 1(0)	+1(0)	-1(0)	- 1(0)	0(0)	
	26	+354(+115)	-347(-112)	-353(-114)	+354(+115)	- 2(-1)	-7(-2)	-1(0)	- 6(-2)	+ 8(+ 3)	
A-D	6	+ 6(+ 6)	- 8(- 8)	- 7(- 7)	+ 10(+ 10)	+ 2(+ 2)	+1(+1)	-2(-2)	- 3(-3)	+ 1(+ 1)	
	17	+ 18(+ 7)	- 12(- 5)	- 20(- 8)	+ 14(+ 6)	- 6(- 2)	+2(+1)	-2(-1)	+ 6(+ 2)	0(0)	
	26	+ 81(+ 26)	- 71(- 23)	- 76(- 25)	+ 69(+ 22)	-10(-3)	-5(-2)	+2(+1)	+ 7(+ 2)	+ 3(+ 1)	
B-D	6	+ 4(+ 4)	- 4(- 4)	- 4(- 4)	+ 4(+ 4)	0(0)	0(0)	0(0)	0(0)	0(0)	
	17	- 29(- 12)	+ 30(+ 12)	+ 30(+ 12)	- 29(- 12)	0(0)	-1(0)	0(0)	- 1(0)	0(0)	
	26	-148(- 48)	+141(+ 46)	+145(+ 47)	-139(- 45)	+ 6(+ 2)	+3(+1)	-2(-1)	- 6(-2)	0(0)	
C-D	6	- 10(- 10)	+ 11(+ 11)	+ 6(+ 6)	- 7(- 7)	- 1(-1)	+4(+4)	-4(-4)	+ 1(+ 1)	0(0)	
	17	+ 43(+ 17)	- 46(- 19)	- 45(- 18)	+ 48(+ 19)	+ 3(+ 1)	+2(+1)	-2(-1)	- 3(-1)	0(0)	
	26	-101(- 33)	+ 93(+ 30)	+104(+ 34)	- 96(- 31)	+ 8(+ 3)	-3(-1)	+3(+1)	- 9(-3)	+ 1(0)	

* Size of sample given in Tables 3, 4, 6, 7 and 8.

frequencies can be obtained from appropriate sums of genotypic frequencies. For example, for loci A and B:

$$\begin{aligned} f(A_1B_1) &= f(A_1B_1, A_1B_1) + f(A_1B_1, A_1B_2) + f(A_1B_1, A_2B_1) + f(A_1B_1, A_2B_2) \\ f(A_1B_2) &= f(A_1B_2, A_1B_1) + f(A_1B_2, A_1B_2) + f(A_1B_2, A_2B_1) + f(A_1B_2, A_2B_2) \\ f(A_2B_1) &= f(A_2B_1, A_1B_1) + f(A_2B_1, A_1B_2) + f(A_2B_1, A_2B_1) + f(A_2B_1, A_2B_2) \\ f(A_2B_2) &= f(A_2B_2, A_1B_1) + f(A_2B_2, A_1B_2) + f(A_2B_2, A_2B_1) + f(A_2B_2, A_2B_2), \end{aligned}$$

and, in general, $f(ij) = \sum_{kl} f(ij, kl)$. Estimation of the gametic frequencies re-

quires that all 10 genotypic frequencies be known but, as mentioned above, the experimental methods of the present study do not separate coupling from repulsion double heterozygotes. However, we see from Tables 3–8 that the total frequency of double heterozygotes is always very low and often zero. Since each $f(ij, kl)$ and $f(il, kj)$, where $i \neq k, j \neq l$, is bounded by zero and the small quantity $1/2f(ikjl)$ we expect good approximation to actual values in our sample if we set $f(ij, kl) = f(il, kj) = (1/2)f(ikjl)$, [e.g. $f(A_1B_1, A_2B_2) = f(A_1B_2, A_2B_1) = 1/4f(A_1A_2B_1B_2)$]. Hence this procedure was followed in computing the frequencies of gametic ditypes. The frequencies of gametic ditypes are not reported since they are very close to frequencies of double homozygotes, given in Tables 3–8. However, we note that highly significant changes in the frequencies of the gametic ditypes occurred in a number of single-generation transitions. In many cases (e.g. A_1B_1 in transition from generation 15 to 16 and A_1C_2 in transition from generation 24 to 25) changes in frequency of 10 standard errors or larger occurred. It is also clear that longer-term changes took place. As examples, the A_1B_2 , A_1C_2 and B_2C_2 ditypes all more than tripled in frequency whereas A_1C_1 and B_1C_1 decreased markedly in frequency over the 22 generation interval studied. Since many of the single generation and also long-term changes in gametic ditypes (and double homozygotes) are larger than can be accounted for by mutation, migration or genetic drift (II) they are evidently due to selection.

Gametic phase disequilibrium: The gametic phase disequilibrium parameter (also called linkage disequilibrium, epistatic disequilibrium, gametic phase unbalance, disequilibrium linkage function and linkage deviation function), Δ , defined as the deviation of observed gametic ditype frequencies from expected frequencies computed as products of allelic frequencies, provides still another measure of the changes which occurred in CCV. For example, gametic phase disequilibrium for loci A and B is given by

$$\begin{aligned} \Delta &= f(A_1B_1) - f(A_1)f(B_1) = f(A_1)f(B_2) - f(A_1B_2) \\ &= f(A_2)f(B_1) - f(A_2B_1) = f(A_2B_2) - f(A_2)f(B_2) \end{aligned}$$

where $f(A_i)$ and $f(B_j)$ are gene frequencies. Gametic phase disequilibrium for other locus pairs is defined by exchanging letters A and B with the letters for such pairs. Values of Δ are given in Table 11. (These values of Δ differ only in third or fourth decimal place from those obtained by setting $f(ij, kl)$ and $f(il, kj)$ equal to zero or to $(1/2)f(ikjl)$; hence we conclude that computing genotypic frequencies on the basis of $f(ij, kl) = f(il, kj) = (1/4)f(ikjl)$ leads to only trivial errors in the

TABLE 11

Measures of gametic phase disequilibrium Δ (top figure), Δ' (middle figure) and r (bottom figure) for six two-locus pairs

Generation	Locus pair					
	A-B	A-C	B-C	A-D	B-D	C-D
Initial*	—0.012	0.011	0.021			
	—0.250	0.055	0.355			
	—0.079	0.045	0.144			
4	—0.025	—0.006	0.028			
	—0.572	—0.041	0.407			
	—0.174	—0.025	0.202			
5	—0.015	—0.023	0.023	0.002	0.001	—0.022
	—0.451	—0.136	0.429	0.009	0.037	—0.112
	—0.112	—0.096	0.173	0.008	0.010	—0.093
6	—0.019	—0.012	0.027	0.008	0.004	—0.008
	—0.354	—0.070	0.376	0.032	0.063	—0.048
	—0.124	—0.050	0.176	0.032	0.024	—0.035
14	—0.047	—0.015	0.042			
	—0.923	—0.128	0.536			
	—0.308	—0.070	0.324			
15	—0.072	—0.046	0.093			
	—0.954	—0.266	0.927			
	—0.395	—0.192	0.531			
16	—0.021	—0.048	0.028	0.008	—0.018	0.011
	—0.459	—0.270	0.483	0.032	—0.377	0.060
	—0.141	—0.200	0.201	0.031	—0.120	0.047
17	—0.019	—0.051	0.034	0.006	—0.012	0.019
	—0.483	—0.303	0.633	0.027	—0.298	0.084
	—0.132	—0.211	0.249	0.026	—0.086	0.076
24	—0.083	—0.061	0.065			
	—0.959	—0.372	0.503			
	—0.413	—0.253	0.318			
25	—0.086	—0.078	0.105	0.016	—0.060	—0.042
	—0.959	—0.381	0.963	0.071	—0.709	—0.217
	—0.421	—0.320	0.503	0.068	—0.297	—0.174
26	—0.098	—0.103	0.114	0.024	—0.047	—0.032
	—0.965	—0.493	0.942	0.146	—0.660	—0.221
	—0.457	—0.422	0.522	0.111	—0.238	—0.144

* Δ , Δ' and r computed from gametic ditype frequencies in the parents of CCV, assuming no change took place during the intercrossing phase of the synthesis of this population. It is unlikely that selection had much effect on Δ during this phase, since propagation was under space planting and survival was high. However, recombination almost certainly led to a reduction in Δ during this phase, so that the values reported in this table are almost certainly overestimates of the actual gametic phase disequilibrium that existed in the initial generation.

estimated values of Δ .) Values of Δ are a function of gene frequencies and hence they have the disadvantage that they cannot be compared directly unless gene frequencies are equal. Values of Δ can, however, be adjusted for gene frequency (LEWONTIN 1965), giving Δ' values which express gametic phase disequilibrium relative to the maximum that Δ can attain [$\Delta' \in (-1.0, 1.0)$]. However, Δ' also has a disadvantage in that it is sensitive to differences in allelic frequency and can

take near maximal values when allelic frequencies differ widely at pairs of loci. The correlation coefficient, r , between alleles is less sensitive to differences in allelic frequencies and has the additional advantage that its sampling distribution is known ($Nr^2 \approx \chi^2$). Consequently we have also reported r values in Table 11. It can be seen from this table that Δ , Δ' and r , although nonzero in the initial generation, were all small and perhaps not larger than expected due to the effects of sampling a small number of parents. Thereafter Δ increased rapidly, until in the latest generations, it had reached more than 90% of its theoretical maximum value for the tightly linked A-B and B-C locus pairs and nearly 50% of its maximum value for tightly linked locus pair A-C. Substantial gametic phase disequilibrium had also developed for the three unlinked locus pairs by the late generations, especially for locus pair B-D for which Δ exceeded 65% of its maximum in generations 25 and 26. The correlation coefficients, r , also indicate the development of substantial gametic phase disequilibrium, especially for the linked pairs of loci and for the B-D combination.

For a pair of adaptively neutral loci linked with amount λ of linkage, it can be shown (WEIR and COCKERHAM, in preparation) that for a population practicing an amount s of selfing, the geometric rate of convergence of Δ to zero soon becomes

$$1/2 \left\{ \frac{1 + \lambda + s}{2} + \left[\left(\frac{1 + \lambda + s}{2} \right)^2 - 2s\lambda \right]^{1/2} \right\}.$$

Thus, for the neutral situation, any original gametic phase disequilibrium that exists will be lost over generations, although in CCV with $t = .0057$ the rate of loss is expected to be low ($< .4$ percent/generation), even for pairs of unlinked loci. In any event the absolute value of Δ is not expected to increase with neutral alleles, as the values in Table 11 are seen to do. Increasing absolute values of Δ have been reported previously in experiments with both outbreeding (e.g. CANNON 1963) and inbreeding species (e.g. HARDING and ALLARD 1968). For random mating populations it can be shown that Δ may increase with tight linkage and selection (discussion in EWENS 1969). Although theory showing conditions under which gametic phase disequilibrium may persist (though not increase) has been restricted in inbreeding populations to adaptively neutral loci (COCKERHAM and WEIR, in preparation), numerical results showing permanent gametic phase disequilibrium in the presence of selection have been reported by JAIN and ALLARD (1966). Analyses of changes in the gametic arrays thus support results of analyses of the zygotic arrays discussed above in indicating striking departures from the neutral situation in CCV. Since the effects of mutation, migration and genetic drift are too small to be measurable in this population (II) the observed changes are evidently due to selection. In the next section we estimate the selective values of the various genotypes to obtain a quantitative measure of the intensity of this selection.

Two-locus selective values: Selective values and their standard errors were computed for the two-locus genotypes according to the methods given in Appendix A. Values obtained for locus pair B-C (Table 12) are given as an example to

TABLE 12

Estimates of two-locus selective values and standard errors (in parentheses) for locus pair B-C

Generation	Genotypes								
	1111	1122	2211	2222	1112	1211	1222	2212	1212
4	0.913 (0.030)	1.262 (0.097)	0.731 (0.190)	1.089 (0.222)	1.345 (0.218)	1.523 (1.127)	0.515 (0.214)	0.826 (0.415)	0.593 (0.509)
5	0.974 (0.034)	1.022 (0.069)	1.787 (0.387)	1.503 (0.252)	0.637 (0.190)	0.000 (0.132)	0.000 (0.037)	0.479 (0.531)	0.000 (0.203)
14	0.881 (0.020)	1.171 (0.092)	0.000 (0.051)	2.041 (0.245)	0.433 (0.093)	0.844 (0.464)	7.439 (1.988)	0.519 (0.209)	2.005 (1.515)
15	0.920 (0.022)	1.531 (0.080)	6.787 (2.357)	0.417 (0.043)	2.995 (0.676)	0.168 (0.410)	0.196 (0.073)	1.838 (0.936)	2.560 (1.479)
16	1.019 (0.025)	1.049 (0.045)	0.715 (0.137)	1.176 (0.130)	0.369 (0.163)	3.244 (4.241)	0.000 (0.054)	0.000 (0.004)	0.000 (0.085)
24	0.954 (0.022)	1.227 (0.046)	0.000 (0.016)	1.550 (0.076)	0.195 (0.063)	1.182 (0.558)	1.091 (0.284)	0.244 (0.071)	0.261 (0.509)
25	0.958 (0.025)	0.936 (0.037)	4.091 (2.480)	1.141 (0.050)	2.404 (0.787)	0.689 (0.583)	0.718 (0.337)	0.580 (0.380)	1.261 (1.993)

illustrate some features of the estimates that were common to all locus pairs. First, the estimates for the heterozygous genotypes have large standard errors and are quite erratic due to sampling errors. The same is the case for the infrequent double homozygotes, e.g. $B_2B_2C_1C_1$. Second, values in successive generations are negatively correlated because they share a common set of genotypic frequencies, used once as $f^{n+1}(x)$ and once as $f^n(x)$ (see II). However, the estimates of selective values of the double homozygotes generally have small standard errors and these estimates reflect the observed trends in gametic and genotypic frequencies. For example, for loci B and C, the increase in $B_2B_2C_2C_2$ at the expense of $B_1B_1C_1C_1$ is reflected by the generally larger selective values of the former. The picture for heterozygotes is obscured by sampling variation.

The selective values for loci B and C are typical in that they do not appear to change in response to changes in genotypic frequencies. For example, the selective values for $B_2B_2C_2C_2$ were much the same in early generations when the B_2C_2 double homozygote was infrequent in the population as in later generations when its frequency had increased several fold. Thus general frequency-dependent selection featuring advantage of infrequent genotypes, i.e., $w(x)$ increasing with decrease in $f(x)$, and vice versa, appears to be unimportant in CCV. On the other hand, nonlinearities that might arise from frequency-dependent selection appear in a number of cases. As an example, the selective values of $B_1B_1C_2C_2$ are higher than those of $B_1B_2C_1C_2$ even though the B_2C_1 gamete is very infrequent in the population. This suggests a compensatory type of frequency-dependent selection. Of course all selective values were estimated from observed genotypic frequencies and so in this sense they are frequency dependent.

Two-locus selective values averaged over generations are given for the six two-locus pairs in Table 13. Estimates of selective values of the homozygotes have

TABLE 13

Estimates of two-locus mean selective values and standard errors (in parentheses)

Locus pair	Genotypes								
	1111	1122	2211	2222	1112	1211	1222	2212	1212
A-B	1.011 (0.015)	1.062 (0.047)	0.969 (0.013)	2.507 (0.919)	1.151 (0.163)	0.985 (0.122)	3.285 (2.157)	1.151 (0.163)	1.031 (0.744)
A-C	0.902** (0.018)	1.266** (0.034)	0.962 (0.017)	1.069 (0.036)	0.964 (0.165)	1.057 (0.181)	1.223 (0.518)	1.448 (0.383)	2.997 (1.777)
B-C	0.946** (0.010)	1.171** (0.027)	2.016 (0.493)	1.274** (0.064)	1.197 (0.156)	1.036 (0.644)	1.423 (0.293)	0.641 (0.176)	0.954 (0.429)
A-D	1.070 (0.038)	0.818* (0.031)	1.171* (0.036)	0.824* (0.029)	4.441 (2.030)	1.287 (0.266)	1.132 (0.269)	3.364 (1.531)	0.000 (6.899)
B-D	1.133* (0.024)	0.809** (0.018)	1.305 (0.124)	1.115 (0.113)	3.141 (1.043)	0.380* (0.133)	0.379 (0.193)	1.458 (0.815)	—†
C-D	1.164* (0.035)	0.799* (0.025)	1.110 (0.043)	0.901 (0.040)	4.378 (1.497)	1.136 (0.195)	1.195 (0.410)	7.296 (7.035)	0.000 (2.902)

* Significant departure from 1.000 at $P < 0.05$.** Significant departure from 1.000 at $P < 0.01$.

† Non-estimable.

small standard errors and reflect the observed trends in genotypic frequencies. Although the selective values of specific heterozygotes are still erratic even when averaged over generations, heterozygotes tend to have higher selective values than double homozygotes. This is reflected in mean selective values of the double homozygotes and the heterozygotes (including double heterozygotes) which are 1.141 and 1.729 respectively. Thus averaged over all of the data of this experiment, the selective values indicate heterozygotes are about 52% superior to homozygotes in reproductive capacity.

In the course of deriving two-locus selection estimates, we also pick up one-locus selective values by taking appropriate marginal totals. For example, we obtain the post-selection one-locus frequencies for A_iA_k as

$$g^n(A_iA_k) \sum_j \sum_l g^n(A_iB_j, A_kB_l)$$

while (Appendix A) pre-selection frequencies are

$$f^n(A_iA_k) \sum_j \sum_l f^n(A_iB_j, A_kB_l),$$

so that the selective values for A_iA_k are given by $w^n(A_iA_k) = g^n(A_iA_k) / f^n(A_iA_k)$. The values for locus B may be obtained similarly. A sample of these marginal values, for loci B and C, are given in Table 14. The single-locus and two-locus estimates bear little relationship to each other in this case, and also for the other locus pairs, as expected considering the complex epistatic nature of the selective forces.

DISCUSSION

For a population undergoing mixed selfing and random outcrossing the condition for the development and maintenance of gametic phase disequilibrium is the

TABLE 14

*Estimates of one- and two-locus selective values and standard errors (in parentheses)
for locus pair B-C in generation 25*

$w(B_1B_1C_1C_1)=0.958$ (0.025)	$w(B_1B_1C_1C_2)=2.404$ (0.787)	$w(B_1B_1C_2C_2)=0.936$ (0.037)	$w(B_1B_1)=0.960$ (0.013)
$w(B_1B_2C_1C_1)=0.689$ (0.583)	$w(B_1B_2C_1C_2)=1.261$ (1.993)	$w(B_1B_2C_2C_2)=0.718$ (0.337)	$w(B_1B_2)=0.817$ (0.192)
$w(B_2B_2C_1C_1)=4.091$ (2.480)	$w(B_2B_2C_1C_2)=0.580$ (0.380)	$w(B_2B_2C_2C_2)=1.141$ (0.050)	$w(B_2B_2)=1.146$ (0.049)
$w(C_1C_1)=0.964$ (0.025)	$w(C_1C_2)=1.640$ (0.313)	$w(C_2C_2)=1.020$ (0.024)	

operation of selection. Pairs of adaptively neutral loci which are in gametic phase equilibrium ($\Delta = 0$) are not expected to develop gametic disequilibrium ($\Delta \neq 0$) (WEIR and COCKERHAM, in press). With adaptive neutrality the frequency of gametic diatypes is also not expected to change in a population such as CCV in which genetic drift, migration and mutation have only trivial effects (II). Yet in CCV highly significant changes occurred in the frequencies of gametic diatypes, and striking gametic phase disequilibrium developed within a few generations for each of the six pairwise combinations of loci which were monitored. These results provide strong evidence that all four of the loci studied, or the linkage blocks they mark, have significant effects on survival.

The data also provide information concerning the complexity of the units on which selection acts in CCV. The A locus, due to the nature of its linkage relationships with Loci B and C, is particularly informative concerning the effects of selection on single loci. These three loci are ordered $B \leftarrow 0.0023 \rightarrow A \leftarrow 0.0048 \rightarrow C$. Since no genes have been reported that occupy a shorter segment of chromosome than that estimated for Locus A (BENZER 1955), it seems likely that no locus other than A occurs between Loci B and C. Thus it is evidently the A locus itself, and not undetected genes linked to this locus, that produces the selective effects associated with the A system. Moreover, these effects are substantial indicating that single enzyme loci can have effects on survival of the same order of magnitude as loci which govern some of the more conspicuous visible polymorphisms.

The data also make it clear that selection operates not only on individual loci but that it also operates differentially on specific two-locus allelic combinations. The development of non-random associations in the gametic and zygotic arrays of both linked and unlinked pairs of loci is evidently attributable to interactions between selection and restriction of recombination resulting from linkage and/or inbreeding. Another feature of selection is that each allele at any one locus interacted favorably with at least one allele and unfavorably with at least one allele at each of the three other loci; thus complicated epistatic interactions occur not only at the two-locus level but, since each locus affects each other locus, also at the three and four locus levels, as shown more precisely by CLEGG, ALLARD and KAHLER (1972). We are therefore lead to the conclusion that the complex behavior of loci A, B, C, and D in CCV results primarily from interactions among selection, linkage and inbreeding.

LITERATURE CITED

- ALLARD, R. W., A. L. KAHLER and B. S. WEIR, 1972 The effect of selection on esterase allozymes in a barley population. *Genetics* **72**: 489-503.
- BENZER, SEYMOUR, 1955 Fine structure of a genetic region in bacteriophage. *Proc. Natl. Acad. Sci. U.S.* **41**: 344-354.
- CANNON, G. B., 1963 The effects of natural selection on linkage disequilibrium and relative fitness in experimental populations of *Drosophila melanogaster*. *Genetics* **48**: 1201-1216.
- CLEGG, M. T., R. W. ALLARD and A. L. KAHLER, 1972 Is the gene the unit of selection?: Evidence from two experimental plant populations. *Proc. Natl. Acad. Sci. U.S.*, in press.
- COCKERHAM, C. CLARK and B. S. WEIR, 1972 Descent measures for two loci with some applications. (In preparation).
- EWENS, W. J., 1969 *Population Genetics*. Methuen and Company, London.
- HARDING, J. and R. W. ALLARD, 1968 Population studies in predominantly self-pollinated species. XII. Interactions between loci affecting fitness in a population of *Phaseolus lunatus*. *Genetics* **61**: 721-736.
- JAIN, S. K. and R. W. ALLARD, 1966 The effects of linkage, epistasis, and inbreeding on population changes under selection. *Genetics* **53**: 633-659.
- KAHLER, A. L. and R. W. ALLARD, 1970 The genetics of isozyme variants in barley. I. Esterases. *Crop Science* **10**: 444-448.
- LEWONTIN, R. C., 1965 The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49-67.
- WEIR, B. S. and C. CLARK COCKERHAM, 1972 Mixed self and random mating at two loci (In preparation).

APPENDIX A

ESTIMATION OF TWO-LOCUS SELECTIVE VALUES

Estimation of two-locus selective values has received little attention, particularly when linkage and inbreeding are taken into account. (The methods given by TURNER 1967, 1968) are not general because his expressions for frequencies of double heterozygotes in inbreeding populations hold only for zero gametic phase disequilibrium (linkage disequilibrium), see COCKERHAM and WEIR, in preparation). Here we extend the maximum likelihood estimators based on genotypic recurrence formulae (ALLARD and WORKMAN 1963) that were used to make the one-locus estimates (II). We develop the theory in some detail to show the necessity for certain assumptions.

The method rests on comparing two sets of genotypic frequencies, one estimated prior to selection and the other estimated after selection has occurred. The selection model we adopt is Model II of WORKMAN and JAIN (1966). Our frequency data are obtained very early in each generation (from seven-day-old seedlings) and so measure genotypic frequencies shortly after the formation of zygotes from gametes of the mature individuals of the previous generation. We assume that the outcrossing which occurs between different individuals in the population is at random. This assumption appears to be reasonable in view of the finding (II) that outcrossing is homogeneous over genotypes in CCV. We also assume that no selection occurs between mating of the mature individuals in one generation and scoring of seedlings in the next generation and, thus, that all selection occurs between scoring and the mating of mature individuals in that generation. This assumption also appears to be reasonable because more than 99% of flowers produced kernels and more than 99% of kernels produced assayable seven-day-old seedlings. Comparing genotypic frequencies $f^n(x)$ of seedlings with genotypic frequencies $g^n(x)$ of mature individuals of genotype x in generation n therefore gives a measure of the relative selective values for this genotype. These selective coefficients are defined as

$$w^n(x) = g^n(x)/f^n(x),$$

so that $f^n(x)$ and $g^n(x)$ may be called the pre- and post-selection frequencies respectively. Under this model, seedling frequencies in generation $n+1$ are determined solely by the mating system and frequencies of mature individuals in generation n . We thus have a functional relationship of the form

$$\{f^{n+1}(x)\} = \Phi(\{g^n(x)\}).$$

By equating these expected $f^{n+1}(x)$ values to their observed values (generation $n+1$ seedling frequencies), the solution to

$$\{g^n(x)\} = \Phi^{-1}(\{f^{n+1}(x)\}),$$

where Φ^{-1} has the observed values substituted, gives the maximum likelihood estimates of the $g^n(x)$. The problem then is to find Φ^{-1} .

We proceed by first displaying the transition equations. Although we use the symbols A and B, these equations clearly apply to any pair of loci linked to an extent λ , and just as clearly to any number of alleles at each locus. For loci A and B we suppose there are alleles A_i and B_j . As before, the frequency of genotypes formed by the union of gametes A_iB_j and A_kB_l is written as $f^n(ij,kl)$ —meaning pre-selection frequencies in generation n . The corresponding post-selection frequency is written as $g^n(ij,kl)$. If we set

$$\begin{aligned} u^n(ij) = & g^n(ij,ij) + \sum_{p \neq i} g^n(ij,pj) + \sum_{q \neq j} g^n(ij,iq) \\ & + \sum_{p \neq i, q \neq j} \left\{ \frac{1+\lambda}{2} g^n(ij,pq) + \frac{1-\lambda}{2} g^n(iq,pj) \right\}, \end{aligned}$$

then the transition equations become:

$$\begin{aligned} f^{n+1}(ij,ij) = & \frac{s}{G} [g^n(ij,ij) + 1/2 \sum_{p \neq i} g^n(ij,pj) + 1/2 \sum_{q \neq j} g^n(ij,iq) \\ & + \sum_{p \neq i, q \neq j} \left\{ \frac{(1+\lambda)^2}{8} g^n(ij,pq) + \frac{(1-\lambda)^2}{8} g^n(iq,pj) \right\}] + \frac{t}{G^2} [u^n(ij)]^2 \\ 2f^{n+1}(ij,kj) = & \frac{s}{G} [g^n(ij,kj) + \frac{1-\lambda^2}{4} \sum_{q \neq j} \{g^n(ij,kq) + g^n(iq,kj)\}] \\ & + \frac{2t}{G^2} [u^n(ij)u^n(kj)] \quad \text{for } k \neq i \\ 2f^{n+1}(ij,il) = & \frac{s}{G} [g^n(ij,il) + \frac{1-\lambda^2}{4} \sum_{p \neq i} \{g^n(ij,pl) + g^n(il,pj)\}] \\ & + \frac{2t}{G^2} [u^n(ij)u^n(il)] \quad \text{for } l \neq j. \\ 2f^{n+1}(ij,kl) = & \frac{s}{G} \left[\frac{(1+\lambda)^2}{4} g^n(ij,kl) + \frac{(1-\lambda)^2}{4} g^n(il,kj) \right] \\ & + \frac{2t}{G^2} [u^n(ij)u^n(kl)] \quad \text{for } k \neq i, l \neq j. \end{aligned}$$

These equations are for double homozygotes, homozygotes at the B locus only, homozygotes at the A locus only, and double heterozygotes, respectively. The divisor $G = \sum_{ijkl} g^n(ij,kl)$ ensures that the pre-selection frequencies in generation $n+1$ sum to one, and so are relative frequencies. If the generation n post-selection frequencies are also to be relative, then $G = 1$ and it will be taken as such. Note that $G = \sum_{ijkl} w^n(ij,kl) f^n(ij,kl)$ also, so that $G = \bar{w}$, the "mean fitness" of generation n . We have imposed unit mean fitness on our population. Another consequence is that $\sum_{ij} u^n(ij) = 1$, so that by adding appropriate equations to obtain the pre-selection gametic frequencies in generation $n+1$:

$$f^{n+1}(A_iB_j) = \sum_{kl} f^{n+1}(A_iB_j, A_kB_l) = su^n(ij) + tu^n(ij) = u^n(ij),$$

so that the quantities $u^n(ij)$ are equal to these gametic frequencies and are thus observable. This leads us to define new functions of the observed generation $n+1$ pre-selection frequencies:

$$h^{n+1}(A_i B_j, A_k B_l) = \frac{1}{s} \{f^{n+1}(A_i B_j, A_k B_l) - t f^{n+1}(A_i B_j) f^{n+1}(A_k B_l)\}, s \neq 0$$

and to write the transition equations in the form:

$$\begin{aligned} h^{n+1}(ij, ij) &= g^n(ij, ij) + 1/2 \sum_{p \neq i} g^n(ij, pi) + 1/2 \sum_{q \neq j} g^n(ij, iq) \\ &\quad + \sum_{p \neq i, q \neq j} \left\{ \frac{(1+\lambda)^2}{8} g^n(ij, pq) + \frac{(1-\lambda)^2}{8} g^n(iq, pj) \right\} \\ 2h^{n+1}(ij, kj) &= g^n(ij, kj) + \frac{1-\lambda^2}{4} \sum_{q \neq j} \{g^n(ij, kq) + g^n(iq, kj)\} \quad k \neq i \\ 2h^{n+1}(ij, il) &= g^n(ij, il) + \frac{1-\lambda^2}{4} \sum_{p \neq i} \{g^n(ij, pl) + g^n(il, pi)\} \quad l \neq j \\ 2h^{n+1}(ij, kl) &= \frac{(1+\lambda)^2}{4} g^n(ij, kl) + \frac{(1-\lambda)^2}{4} g^n(il, kj) \quad k \neq i, l \neq j. \end{aligned}$$

This system of equations, which may be summarized as

$$\{h^{n+1}(x)\} = \Theta \{g^n(x)\},$$

has the advantage over the system Φ of being linear, and hence easy to transform to the system

$$\{g^n(x)\} = \Theta^{-1} \{h^{n+1}(x)\}.$$

The equations of this inverse system follow, where the $h^{n+1}(x)$ refer to their observed values, so that the $g^n(x)$ are maximum likelihood estimates (BAILEY 1951). Note that for free recombination ($\lambda=0$) both double heterozygotes for any four alleles have the same coefficient in the transition equations, so that we cannot find both of $g^n(ij, kl)$ and $g^n(il, kj)$ but only their sum.

$\lambda \neq 0$

$$\begin{aligned} g^n(ij, ij) &\hat{=} h^{n+1}(ij, ij) - \sum_{p \neq i} h^{n+1}(ij, pi) - \sum_{q \neq j} h^{n+1}(ij, iq) \\ &\quad + \frac{1}{1+\lambda^2} \sum_{p \neq i, q \neq j} \{-2\lambda^2 h^{n+1}(ij, pq) + (1-\lambda^2) h^{n+1}(iq, pj)\} \\ g^n(ij, kj) &\hat{=} 2h^{n+1}(ij, kj) - \frac{1-\lambda^2}{1+\lambda^2} \sum_{q \neq j} \{h^{n+1}(ij, kq) + h^{n+1}(iq, kj)\} \quad k \neq i \\ g^n(ij, il) &\hat{=} 2h^{n+1}(ij, il) - \frac{1-\lambda^2}{1+\lambda^2} \sum_{p \neq i} \{h^{n+1}(ij, pl) + h^{n+1}(il, pi)\} \quad l \neq j \\ g^n(ij, kl) &\hat{=} \frac{1}{\lambda(1+\lambda^2)} \{(1+\lambda)^2 h^{n+1}(ij, kl) - (1-\lambda)^2 h^{n+1}(il, kj)\} \quad k \neq i, l \neq j. \end{aligned}$$

$\lambda = 0$

$$\begin{aligned} g^n(ij, ij) &\hat{=} h^{n+1}(ij, ij) - \sum_{p \neq i} h^{n+1}(ij, pi) - \sum_{q \neq j} h^{n+1}(ij, iq) \\ &\quad + 1/2 \sum_{p \neq i, q \neq j} \{h^{n+1}(ij, pq) + h^{n+1}(iq, pi)\} \\ g^n(ij, kj) &\hat{=} 2h^{n+1}(ij, kj) - \sum_{q \neq j} \{h^{n+1}(ij, kq) + h^{n+1}(iq, kj)\} \quad k \neq i \\ g^n(ij, il) &\hat{=} 2h^{n+1}(ij, il) - \sum_{p \neq i} \{h^{n+1}(ij, pl) + h^{n+1}(il, pi)\} \quad l \neq j \\ \{g^n(ij, kl) + g^n(il, kj)\} &\hat{=} 4\{h^{n+1}(ij, kl) + h^{n+1}(il, kj)\} \quad k \neq i, l \neq j. \end{aligned}$$

For the case of two alleles per locus, there are ten possible genotypes. When all ten pre-

selection frequencies can be determined, we can estimate all ten post-selection frequencies in the previous generation (although they are constrained to sum to one). In the present case there are just nine observable genotypes, and while we could still estimate ten post-selection frequencies we note that the two double heterozygote frequencies are constrained to have a difference depending on the amount of gametic phase disequilibrium. We thus present just nine post-selection frequencies also—by taking the sum:

$$\{g^n(ij,kl) + g^n(il,kj)\} = \frac{4}{1+\lambda^2} \{h^{n+1}(ij,kl) + h^{n+1}(il,kj)\} \text{ for all } \lambda.$$

We stress, however, that setting both pre-selection double heterozygote frequencies equal to each other neither requires nor implies that corresponding post-selection frequencies, and hence selective values, are equal unless gametic phase equilibrium prevails.

Once the $g^n(x)$ are estimated, we need only divide by the corresponding $f^n(x)$ in order to estimate the selective value for the x^{th} genotype. Should the value of some $f^n(x)$ be zero, the selective value is not estimable. Should a $g^n(x)$ be negative (as will happen if fewer of that genotype are observed in generation $n+1$ than the amount of outcrossing alone would require), we say that the selective value of that genotype is zero. Both of these situations occurred occasionally for heterozygotes, presumably as a result of sampling error.

The selection coefficients $w^n(x)$ are functions of $f^n(x)$, $f^{n+1}(x)$, t and λ . The estimates we have of all these quantities are subject to sampling error, and we can give estimates of the variance of each of the estimates (see Appendix of II). We can, therefore, estimate the variance of $w^n(x)$ from these variances by using large sample theory (CRAMÉR, 1946). To the order of accuracy of each of the estimates of $f^n(x)$, $f^{n+1}(x)$, t and λ :

$$\text{var}(w^n(x)) = \frac{1}{[f^n(x)]^2} \text{var}(g^n(x)) + \frac{[g^n(x)]^2}{[f^n(x)]^4} \text{var}(f^n(x)),$$

where

$$\begin{aligned} \text{var}(g^n(x)) = & \sum_{yz} \frac{\partial g^n(x)}{\partial f^{n+1}(y)} \frac{\partial g^n(x)}{\partial f^{n+1}(z)} \text{cov}(f^{n+1}(y), f^{n+1}(z)) \\ & + \left[\frac{\partial g^n(x)}{\partial \lambda} \right]^2 \text{var}(\lambda) + \left[\frac{\partial g^n(x)}{\partial t} \right]^2 \text{var}(t), \end{aligned}$$

and all quantities are taken at their estimated or observed values.

An alternative method of treating the case of two alleles at each of two tightly-linked loci is to assume that the two pairs of alleles represent four alleles at a single locus. This treatment for locus pairs A-B, A-C and B-C gave selective values within 0.02 of those found as above. The mechanics of obtaining such estimates have been described previously (II).

LITERATURE CITED

- ALLARD, R. W. and P. L. WORKMAN, 1963 Population studies in predominantly self-pollinated species. IV. Seasonal fluctuations in estimated values of genetic parameters in lima bean populations. *Evolution* **17**: 470-480.
- BAILEY, N. T. J., 1951 Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics* **7**: 268-274.
- CRAMÉR, H., 1946 *Mathematical methods of statistics*. Princeton.
- TURNER, J. R. G., 1967 On supergenes: I. The evolution of supergenes. *Am. Naturalist* **107**: 195-221. —, 1968 On supergenes: II. The estimation of gametic excess in populations. *Genetica* **39**: 82-93.
- WORKMAN, P. L. and S. K. JAIN, 1966 Zygotic selection under mixed random mating and self fertilization: theory and problems of estimation. *Genetics* **54**: 159-171.